Cognition 204 (2020) 104373

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Patterns of bilingual language use and response inhibition: A test of the adaptive control hypothesis

Patrycja Kałamała[a,*], Jakub Szewczyk[a,b], Adam Chuderski[c], Magdalena Senderecka[c], Zofia Wodniecka[a,*]

[a] Psychology of Language and Bilingualism Lab, Institute of Psychology, Jagiellonian University in Krakow, Ingardena 6, 30-060 Krakow, Poland
[b] Department of Psychology, University of Illinois, Urbana-Champaign, United States of America
[c] Institute of Philosophy, Jagiellonian University in Krakow, Grodzka 52, 31-044 Krakow, Poland

ARTICLE INFO

Keywords:
Bilingualism
Bilingual advantage
Inhibition
Executive functions
Factor analysis
SEM

ABSTRACT

Given prior studies that provided inconsistent results, there is an ongoing debate on the issue of whether bilingualism benefits cognitive control. We tested the Adaptive Control Hypothesis, according to which only the intense use of different languages in the same situation without mixing them in single utterances (called dual-language context) confers a bilingual advantage in response inhibition.

In a large-scale correlational study, we attempted to circumvent several pitfalls of previous research on the bilingual advantage by testing a relatively large sample of participants and employing a more reliable and valid measurement of constructs (i.e., latent variable approach accompanied by Bayesian estimation). Our results do not support the Adaptive Control Hypothesis' prediction: the intensity of the dual-language context experience was unrelated to the efficiency of response inhibition in bilinguals.

The results suggest that the Adaptive Control Hypothesis is not likely to account for the inconsistent results regarding the bilingual advantage hypothesis, at least in the case of the response-inhibition mechanism. At the same time, the study points to the problem of measuring the response-inhibition construct at the behavioral level. No evidence for a robust response-inhibition construct adds to the growing skepticism on this issue in the literature.

1. Introduction

The bilingual advantage hypothesis posits that the experience of managing two languages trains and enhances bilinguals' cognitive control (i.e., the set of cognitive processes responsible for goal-directed behavior; for an overview see Friedman, 2016; Kroll, Bobb, & Hoshino, 2014). Over the last 20 years, many studies have tested the bilingual advantage hypothesis but no consensus has been reached (for meta-analyses and reviews see Donnelly, Brooks, & Homer, 2015; Paap, 2019; van den Noort et al., 2019). On the one hand, bilinguals outperform monolinguals on a range of experimental tasks that tap into different aspects of cognitive control, such as attentional control, inhibition, and task switching. On the other hand, differences between bilinguals and monolinguals cannot be consistently replicated, especially in large-scale studies (e.g., Dick et al., 2019; Paap & Greenberg, 2013; von Bastian, Souza, & Gade, 2016). Given these conflicting findings, the bilingual advantage hypothesis has been extensively debated and seriously questioned (see the discussion article by Paap, Johnson, & Sawi, 2015, and the corresponding commentaries in *Cortex*, 2015, vol. 73). The debate led researchers to argue that if the bilingual advantage exists at all, it is not as robust as previously assumed and might be restricted to specific groups of individuals and/or specific cognitive processes (Bak, 2015, 2016a, 2016b; Bialystok, 2016; de Bruin, 2019). Recently, there have been explicit calls for a revision of this hypothesis (Blanco-Elorrieta & Pylkkänen, 2018; Dick et al., 2019; von Bastian et al., 2016). In light of this, researchers have started to investigate which aspects of bilingual language use might be responsible for shaping cognitive control efficiency. While the majority of studies focused on the role of language switching (e.g., Beatty-Martínez & Dussias, 2017; Jylkkä et al., 2017; Verreyt, Woumans, Vandelanotte, Szmalec, & Duyck, 2016), there are several that took into account more aspects related to bilingual language experience (e.g., L2 proficiency, language dominance and code-switching in Samuel, Roehr-Brackin, Pak, & Kim, 2018; proficiency in foreign languages and amount of use of foreign languages

* Corresponding authors.
E-mail addresses: patrycja.kalamala@uj.edu.pl (P. Kałamała), adam.chuderski@uj.edu.pl (A. Chuderski), magdalena.senderecka@uj.edu.pl (M. Senderecka), zofia.wodniecka@uj.edu.pl (Z. Wodniecka).

https://doi.org/10.1016/j.cognition.2020.104373
Received 4 October 2019; Received in revised form 27 April 2020; Accepted 7 June 2020
Available online 22 June 2020

in Boumeester, Michel, & Fyndanis, 2019).

A recent theoretical framework for investigating how differences in language experience of bilinguals affect cognitive control has been proposed as the Adaptive Control Hypothesis (Green & Abutalebi, 2013; ACH hereafter). The ACH posits that bilinguals differ in the engagement of cognitive control depending on the patterns in which they use their languages in everyday communication (the so-called interactional context in the ACH; pp. 515–516 in the paper). In the current study, we took a comprehensive approach to assess the patterns of language use in bilinguals (as proposed in the ACH) and investigated their relationship with the efficiency of cognitive control. The study thereby has the potential to shed light on the circumstances under which cognitive advantages of bilingualism can be expected and explain at least some of existing inconsistencies in the literature.

1.1. The adaptive control hypothesis and evidence supporting it

The ACH distinguishes three main patterns of everyday language use in bilinguals: single-language context, dual-language context, and dense-code switching. Individuals that represent the single-language context (SLC) speak only one language in each context (e.g., one language at home, another one at work), while individuals that represent either the dual-language context or dense-code switching (DLC or DCS, respectively) speak two languages within the same context (e.g., at home and/or at work). However, when two languages are spoken in the same context (DLC and DCS), the pattern of language use can still vary. DLC occurs when distinct languages are spoken with distinct speakers (i.e. languages are not mixed in one and the same utterance), whereas DCS occurs when elements of languages are freely mixed in single utterances.

Based on the distinction proposed by the ACH, the patterns of language use in bilinguals can be described in two dimensions: (1) *co-occurrence of languages* (i.e., the extent to which languages are used within the same context) and (2) *frequency of language mixing* (i.e., the extent to which linguistic elements of languages, e.g. words, are mixed within single utterances). The first dimension makes it possible to differentiate between bilinguals who use different languages in different contexts (i.e. SLC bilinguals) and those who use more than two languages within the same context (i.e. either DLC bilinguals or DCS bilinguals). The second dimension makes it possible to differentiate between bilinguals who mix their languages (i.e. DCS bilinguals) and those who do not (i.e. either SLC or DLC bilinguals).

According to the ACH, effective communication of bilinguals involves the workings of cognitive control. Crucially, different patterns of language use engage different cognitive processes and only those processes that are actively involved in effective communication are trained and show greater efficiency. Table 1 presents the main differences between bilinguals' patterns of language use and their relations to specific cognitive processes.

As shown in Table 1, DCS bilinguals virtually do not engage cognitive control, because they mix their languages freely, and therefore do not require control processes to oversee the currently used language. In contrast, bilinguals who operate in either SLC or DLC need to engage key cognitive processes in language use, because they have to constrain the currently used language. Crucially, some cognitive processes are engaged in both SLC and DLC, but there are also some that are engaged only in the case of DLC. This distinction is particularly evident in cognitive processes related to inhibition (see Table 1). The ACH indicates two inhibition processes, i.e., *interference control* (the ability to ignore distracting information) and *selective response inhibition* (the ability to suppress a dominant or on-going response). Both SLC and DLC bilinguals require efficient interference control in order to suppress spurious activations of the language currently not in use. At the same time, only DLC bilinguals require efficient response inhibition which allows inhibiting a currently used language and switching to another one during a conversation.

Although the ACH offers an interesting insight into the mechanisms underlying the bilingual advantage in different cognitive processes, to the best of our knowledge only a few studies have been directly dedicated to testing the predictions of the ACH in healthy adults (Beatty-Martínez et al., 2019; Hartanto & Yang, 2016, 2020; Henrard & Van Daele, 2017; Ooi, Goh, Sorace, & Bak, 2018; Pot, Keijzer, & de Bot, 2018). Some of these studies provided evidence in favor of the ACH. In two consecutive experiments, Hartanto and Yang (2016, 2020) showed that DLC bilinguals (compared to SLC bilinguals) demonstrate more efficient task switching, as indicated by their smaller switch costs in reaction times (RTs). Similarly, Ooi et al. (2018) reported that DLC bilinguals (compared to SLC bilinguals) exhibit more efficient interference control, as reflected by their smaller congruency effect in RTs in the flanker task (for similar findings see Beatty-Martínez et al., 2019; Pot et al., 2018). However, there are also findings that contradict the predictions of the ACH. For example, Pot et al. (2018) did not find support for the relationship between the diversity of language use (indicating SL context) and task switching (as assessed by error rate in the Wisconsin Card Sorting Task), while Hartanto and Yang (2020) showed that better goal maintenance (as assessed by mixing costs in RTs) and interference control (as assessed by flanker effects in RTs) were related to greater exposure to DCS but unrelated to exposure to DLC and SLC, which is at odds with the predictions of the ACH. Taken together, the available evidence for the ACH is inconsistent. While some studies reported effects in favor of the ACH (e.g., Hartanto & Yang, 2016, 2020, for task switching; Pot et al., 2018, for interference control; Beatty-Martínez et al., 2019, for proactive control), others did not find supportive evidence (e.g., Hartanto & Yang, 2020, for interference control and goal maintenance; Pot et al., 2018, for switching). Moreover, the available evidence mostly concerns the efficiency of task switching and interference control, while the ACH also distinguishes other cognitive constructs, e.g., response inhibition, that have not yet been thoroughly examined (see also Appendix A). The current study aims to fill this gap.

Table 1
Overview of the Adaptive Control Hypothesis. Differences between bilinguals' patterns of language use (top) and expected efficiency of cognitive processes relative to monolingual speakers in a monolingual environment (bottom) based on the Adaptive Control Hypothesis.

	Pattern of language use		
	Single-language	Dual-language	Dense-code switching
Dimension			
Co-occurrence of languages: *Do languages co-occur in a context?*	No	Yes	Yes
Language mixing: *Are languages mixed in one and the same utterance?*	No	No	Yes
Cognitive process			
Opportunistic planning	–	–	✓
Goal maintenance, interference control	✓	✓	–
Salient cue detection, selective response inhibition, task engagement-disengagement	–	✓	–

Notes. ✓, bilingual advantage expected; –, no bilingual advantage expected.

1.2. Methodological limitations of studies testing the bilingual advantage hypothesis

As discussed above, according to the ACH, the cognitive benefits of bilingualism are determined by patterns of language use (interactional context). Therefore, the ACH provides an interesting explanation of why the cognitive advantages of bilingualism are not consistently observed across studies. However, this mixed evidence may also be driven by methodological shortcomings in many of the published studies that tested the bilingual advantage hypothesis (for arguments see Bakker, 2015; Dick et al., 2019; Linck, 2015; Marzecová, 2015; Paap et al., 2015; von Bastian et al., 2016). As shown in recent reviews (Donnelly et al., 2015; Paap, 2019; van den Noort et al., 2019), previous studies frequently tested small samples of participants, thus leading to imprecise measurements and low statistical power. In addition, many of them did not adequately control for potential confounds such as age, intelligence and socioeconomic status, which have been shown to affect cognitive control abilities (Antoniou, 2019; Bak, 2016a; Samuel et al., 2018). Based on these methodological shortcomings, some researchers have argued that results supporting the bilingual advantage might be either Type-II errors or effects of factors other than bilingualism (Bakker, 2015; Paap et al., 2015; von Bastian et al., 2016).

In addition to the issues listed above, there are also another methodological problems with previous research that are related to the reliability and validity of the experimental tools used to investigate the relationship between bilingualism and cognitive control. First of all, most of the previous studies did not report the psychometric properties of cognitive control tasks (Friedman, 2016; Paap & Sawi, 2016), which suggests that researchers assumed that these tasks have sufficient psychometric properties to detect the true variance related to a task manipulation. However, as studies testing the psychometric characteristics of cognitive control tasks suggest, this is unlikely to be the case (Hedge, Powell, & Sumner, 2017; Paap & Sawi, 2016; Parsons, Kruijt, & Fox, 2018). The reliability of a task (informing about the true variance reflected by a task) is not stable for a given experimental paradigm but depends on the specificity of a task procedure (e.g., high vs. low number of trials) and the characteristics of the participants' sample (e.g., participants with high vs. low cognitive abilities). Therefore, if a task is unreliable in a study, it could either mask the true effect or produce a false one. Second, most of the previous studies on the bilingual advantage employed only a single experimental task, following the assumption that single tasks have sufficient psychometric properties to detect inter-individual variation within the cognitive process targeted (for descriptions of classic tasks and their links to cognitive processes in bilinguals, see Valian, 2015a, 2015b). However, variance in single-task measures is driven not only by the cognitive process in question but also by other irrelevant factors, such as perceptual processing of stimuli, memorizing the task rules, and motivation. In consequence, the measurement of the targeted process – even if reliable – can be contaminated by other task-specific processes. These challenges in accurate measurement of cognitive control, referred to as the task impurity problem, were raised two decades ago (Miyake et al., 2000) and are currently hotly debated in cognitive science literature (e.g., Friedman & Miyake, 2017; Haaf & Rouder, 2019; Karr et al., 2018).

Furthermore, the psychometric problems mentioned above pertain not only to cognitive control tasks, but also to the assessment of bilingualism in previous studies. Available evidence for the relationship between bilingualism and cognitive control mostly comes from research that assessed bilinguals' language experience using questionnaires. However, many studies have not reported the psychometric characteristics of their questionnaire-derived measures, even if a completely new questionnaire was proposed (e.g., Hartanto & Yang, 2016, 2020; Pot et al., 2018; cf., Rodriguez-Fornells, Krämer, Lorenzo-Seva, Festman, & Münte, 2012). Therefore, there seems to be an underlying assumption that questionnaires provide reliable assessment of language users' experience; however, as we will show below, this is not necessarily the case.

Because the study by Hartanto and Yang (2016) provided the first empirical evidence in favor of the ACH, it is often considered as key evidence for the role of patterns of language use in shaping the cognitive advantages of bilingualism (e.g., Barbu, Orban, Gillet, & Poncelet, 2018; Bialystok, 2017; Blanco-Elorrieta & Pylkkänen, 2018; de Bruin, 2019). In this study, Hartanto and Yang assessed bilingual language use by means of a questionnaire developed for the purpose of this experiment. However, since the authors did not report the psychometric characteristics of their questionnaire-derived measures, it is not clear whether their measurement was reliable. Thanks to the fact that the authors shared the dataset used in their study online, we were able to perform a reanalysis of their data (for detailed description and outcomes see Appendix B). The reanalysis showed that only one out of the six questionnaire-derived measures (the composite score of DLC bilingualism) produced statistically significant effects in this study; however, this measure had poor psychometric value (i.e., questionnaire items weakly correlated with each other, for details see Appendix B). In consequence, the poor reliability of the critical measure raises a concern as to whether the study by Hartanto and Yang (2016) indeed provides compelling arguments for the ACH.

Interestingly, the problems of reliable and valid measurement presented above can be overcome by employing the latent variable approach (Friedman, 2016; Gade, 2015; Kline, 1998). In this approach, several measures assumed to tap into the same theoretical construct are employed and the common variance among them is extracted. This common variance, referred to as the latent variable (or factor), assesses whether and to what extent different measures tap into the same theoretical construct. It also provides direct information about the reliability of the measures. The properties of this analysis could therefore be useful in testing the relationship between bilingualism and cognitive control. Specifically, when different measures share variance but the effect is observed in only one of them, it is likely that this effect is either spurious or related to task- or item-specific variance. In such a case, there would be no basis for a conclusion about the general bilingual advantage. Although the latent variable approach can provide a reliable and valid measurement of the targeted constructs related to bilingualism and cognitive control, to the best of our knowledge there have been only a few attempts to use this approach when testing the cognitive effects of bilingualism (e.g., Hartanto & Yang, 2020; Jaekel, Jaekel, Willard, & Leyendecker, 2019). Importantly, however, the use of this approach in previous research was limited to the assessment of cognitive control constructs, but none of the studies so far adopted this approach to assess bilingualism and its relationship with cognitive control in one study.

1.3. The present study

In the current study, we investigated the relationship between patterns of language use and the efficiency of cognitive control in bilinguals. Our goal was to test one of the predictions of the ACH while avoiding the pitfalls of previous research on the bilingual advantage (i.e., small samples of participants, the task impurity problem). We focused on the process of response inhibition (i.e., the ability to suppress a dominant or ongoing response), which is often associated with language control in bilinguals but has not been well studied in the context of the ACH (but see Henrard & Van Daele, 2017). Drawing on the ACH (Green & Abutalebi, 2013), we predicted that the higher the intensity of a DLC experience (i.e., the extent to which different languages co-occur and are not mixed within one and the same utterance in everyday communication), the better bilinguals' response inhibition.

To address the measurement problems discussed in Section 1.2, we employed the latent variable approach, which should allow a more reliable and more valid measurement of bilingualism and cognitive control, as compared to previous studies testing the cognitive effects of bilingualism. The intensity of DLC experience was defined as the

interplay between the co-occurrence of languages and the frequency of language mixing (for definitions see Section 1.1.). This was assessed using two questionnaires. The measurement of response inhibition was guided by previous latent-variable work (Chuderski, Taraday, Nęcka, & Smoleń, 2012; Friedman & Miyake, 2004; Karr et al., 2018; Miyake et al., 2000). Crucially, Friedman and Miyake (2004) proposed a model of inhibition in which one of the factors, i.e., inhibition of prepotent response (defined as the ability to deliberately suppress dominant, automatic, or prepotent responses), is conceptually similar to the construct of response inhibition proposed in the ACH. In order to measure response inhibition, we selected four classic tasks shown to form the response-inhibition factor in previous work (Chuderski et al., 2012; Friedman & Miyake, 2004; Karr et al., 2018): the antisaccade task, the go/no-go task, the Stroop task and the stop-signal task. Importantly, two of these tasks (the go/no-go task and the Stroop task) were also proposed by the ACH to test response-inhibition skills in bilinguals (Green & Abutalebi, 2013, pp. 522–523).

The battery of tasks and questionnaires was administered to a relatively large group of Polish-English adult bilinguals who declared the use of two languages on a daily basis. First, we tested whether the psychometric properties of the measures enabled identification of the two latent variables in question, namely DLC intensity and response inhibition. After having identified the respective latent variables, the ACH's prediction was tested using Structural Equation Modeling. We built a model testing the relationship between DLC intensity and response inhibition while controlling for covariates that could confound this relationship, namely socio-economic status and intelligence.

2. Method

2.1. Participants

Two hundred and fifteen participants took part in the study. Three participants were excluded due to technical problems with data collection. Another three were excluded because they were considerably older than the other participants (i.e., over 55 years old) and might have suffered from age-related deterioration of cognitive functioning (Rey-Mermet, Gade, & Oberauer, 2018). In addition, fourteen participants were excluded for the following reasons: Polish acquired as a third or later language (3 participants); incorrect completion of the questionnaires on language experience (i.e., Polish not included as one of the languages; 5 participants); incorrect performance of the response-inhibition tasks included in the main analyses (i.e., >90% responses incorrect in either the antisaccade task, the go/no-go task, or the Stroop task; 6 participants). In addition, nine participants who incorrectly performed the stop-signal task (i.e., had accuracy on trials including stop-signal higher/lower than 90%/10% and/or stop-signal reaction time shorter than 50 ms; following the exclusion criteria by Congdon et al., 2012) were removed from the stop-signal task but were included in the overall dataset as finally this task did not enter the main analyses (for a justification see Section 3.2.). In total, one hundred and ninety-five participants were included in the reported analyses (mean age 24 years, 147 female, 99 right-handed).

The participants' fluid intelligence was measured using a shortened version of Raven's Advanced Progressive Matrices test (only odd-numbered items, 20 min to complete; score as the sum of correct responses). On average, participants scored 71% on this test. The participants' socio-demographic characteristics was assessed using a socio-demographic background questionnaire. Overall, participants and their parents were well-educated (equivalent to “high school completed”), considered themselves as having relatively high social status, and had a moderate to high income (for details, see Table 2). Almost all participants reported everyday computer use; around one-quarter of them reported playing computer games; 17 participants played musical instruments on a daily basis.

Data on the participants' language proficiency and history of language learning were collected using a language background questionnaire based on Li, Zhang, Tsai, and Puls (2014) and Marian, Blumenfeld, and Kaushanskaya (2007). Participants were language-unbalanced Polish-English bilinguals whose only language acquired in early childhood was Polish (L1). On average, they started learning English (L2) in primary school at around the age of seven. They started using English more intensively when they attended junior high school at around the age of 12. Table 3 presents self-assessment data concerning the participants' language abilities. Participants considered their L1 proficiency to be significantly higher than their L2 proficiency, which they considered intermediate to high. At the same time, their overall L2 proficiency was substantially higher than that of their additional languages. On average, they used L1 for slightly over half the day and 35% of them more frequently used L2 than L1 on any given day. In addition, about 30% of them declared the use of some additional languages (predominantly German, Spanish or French). However, the overall proficiency of these additional languages was relatively low and the participants used these languages only occasionally. In order to obtain an objective measure of L2 proficiency, the participants' vocabulary knowledge of English was assessed using the LexTALE test (participants decided whether a string of letters is a correct English word or not; Lemhöfer & Broersma, 2012). On average, participants scored relatively high (a mean of 79 points out of 100, *SD* = 10), thus indicating that they were moderately to highly proficient in L2. As such, the results of the LexTALE test are consistent with the self-rated proficiency.

Table 2
Participants' socio-demographics based on a self-assessment questionnaire.

Statistic	*N*	*Mean*	*SD*	Min	Max
Age (in years)	195	24.13	4.72	19	42
Fluid intelligence test score (1–18)[a]	195	12.81	2.63	7	18
Education (1–4)[b]	195	2.44	0.52	2	4
Parental education (1–4)[b]	195	2.39	0.68	1	4
Social status (1 − 10)[c]	195	6.26	1.33	3	9
Income (1–6)[d]	195	2.91	1.18	0	6
Computer use (hours per day)	192	5.15	2.73	1.00	16.00
Playing computer games (hours per day)	48	1.62	1.11	0.50	1.50
Playing a musical instrument (hours per day)	17	4.76	3.58	1.00	12.00

Notes. *SD*, standard deviation.
[a] Score in a version of Raven's Advanced Progressive Matrices test (for description see the text).
[b] Self-ratings were 1 = less than high school to 4 = more than Master degree.
[c] Self-ratings were 1 = people who have the least money, the least education, and the least prestigious jobs or no job to 10 = people who have the most money, the most education, and the most prestigious jobs.
[d] Self-ratings were 1 = <500 EUR per month to 6 = >4500 EUR per month.

The study met the requirements and gained the approval of the Ethics Committee of Jagiellonian University in Krakow, Institute of Psychology, concerning empirical studies with human participants. Polish-English bilinguals were recruited via a job-hunting internet portal. Only individuals that declared the everyday use of both Polish and English were invited to participate in the study. The participants were not aware of the reasoning behind the study. Instead, they were told that the study concerns the effectiveness of cognitive and language abilities. Each participant signed an informed consent form prior to the procedure. Following the testing, the participants were debriefed, informed about the goals of the study and paid for their participation (PLN 40, about $10).

2.2. Measures

The battery consisted of four questionnaires, four response-inhibition tasks, two linguistic tasks (i.e., the LexTALE test, and a semantic

Table 3
Participants' language experience based on self-assessment questionnaires.

	Polish (L1) $N = 195$				English (L2) $N = 195$				Additional languages (L3-L5)[a] $N = 66$			
Statistic	*Mean*	*SD*	Min	Max	*Mean*	*SD*	Min	Max	*Mean*	*SD*	Min	Max
Overall proficiency	8.90	0.48	4.25	9.00	7.84	0.87	4.50	9.00	5.55	1.84	2.00	9.00
Listening	8.95	0.37	5	9	7.89	0.88	5	9	5.76	1.87	2	9
Reading	8.92	0.60	2	9	8.03	0.86	5	9	6.05	1.87	2	9
Speaking	8.90	0.52	5	9	7.78	1.11	3	9	5.10	2.08	1	9
Writing	8.82	0.72	3	9	7.67	1.19	3	9	5.29	2.07	2	9
Age of acquisition[b]	0.06	0.46	0	5	6.65	3.19	0	26	13.79	6.02	0	29
Age of active use [b]	1.16	2.09	0	12	11.74	5.05	0	35	17.60	5.79	0	37
% of daily use	54	18	4	96	42	18	4	96	11	9	1	39

Notes. *SD*, standard deviation; the self-ratings range for proficiency was from 1 = no knowledge of given language to 9 = native-like proficiency.
[a] Statistics for the average of L3–L5.
[b] Age in years.

written fluency task), and the paper-and-pencil intelligence test (i.e., a version of Raven's Advanced Progressive Matrices test, as described in Section 2.1). The questionnaires included the socio-demographic background questionnaire, the language background questionnaire, and two questionnaires assessing the patterns of language use: Patterns of Language Use Questionnaire developed in this study and Code-switching and interactional contexts questionnaire published in Hartanto and Yang (2016). The questionnaires were administered using electronic PDF forms and the tasks were administered using DMDX (Forster & Forster, 2003).

2.2.1. Measurement of dual-language context intensity

The intensity of DLC experience in bilinguals (i.e., the intensity of using different languages in the same context without mixing them in single utterances) was assessed via two questionnaires: The Patterns of Language Use Questionnaire and The Code-switching and Interactional Contexts Questionnaire. For each questionnaire, we calculated an individual measure of DLC intensity (DLC intensity 1 and DLC intensity 2, respectively). The following subsections describe the computation of the DLC intensity measures in each questionnaire. The exemplary data and the R script needed to reproduce the measures are available at https://osf.io/5x9sc/.

2.2.1.1. Patterns of language use questionnaire (DLC intensity 1)

2.2.1.1.1. Overview. The questionnaire consists of four parts, each of which targets a different social context: home, work, school or free time (see Appendix C). For each context, participants list all languages they use in this context and estimate how many hours a day they use these languages in that context. If they use more than one language in a given context, they additionally assess the extent to which the four statements adequately represent their language-use habits in that context. The statements describe different everyday situations of language mixing that are typical for Polish-English bilinguals. These statements are the same for each context and are accompanied by examples of the situations specific to a given context. The statements are assessed on a scale from 1 (never) to 9 (always).[1]

2.2.1.1.2. Measures. DLC intensity 1 consisted of two component measures: language entropy and language mixing. Language entropy served to assess the co-occurrence of languages and was computed using Shannon's entropy formula. Shannon's entropy corresponds to uncertainty about the use of a language; in the present study, it characterizes the probability distribution of all languages used in a given context. When many languages are equally likely to be used, the entropy is high; when only a single language is likely, the entropy is zero. The entropy was computed for each context separately using the following formula:

$$\sum_{i=1}^{n} - p_i log_2 p_i$$

where p_i is the probability of the use of a language in a context (see also Gullifer & Titone, 2018). In order to obtain the overall language entropy, the entropies were averaged across the contexts with regard to the amount of time spent on using any language within these contexts. A higher score indicates more balanced use of different languages during a day.

Language mixing served to assess how frequently elements of two languages (e.g. words) were mixed in single utterances and was computed based on the self-rated habits of language use (statements 2–4).[2] The responses were first averaged within contexts and then averaged across contexts with weights proportional to the time spent using more than one language in each context. A lower score indicates less frequent mixing of languages within the same utterance.

DLC intensity 1 was computed by summing the standardized values of the language entropy and the standardized reversed values of the language mixing for each participant. A higher score of DLC intensity 1 indicates a greater intensity of DLC in a bilingual.

2.2.1.2. Code-switching and interactional contexts questionnaire

2.2.1.2.1. Overview. The questionnaire is a Polish translation of the questionnaire used by Hartanto and Yang (2016; see Appendix E in their paper). The only modification in comparison to the original questionnaire was the use of a 9-point scale (1, never, to 9, always) instead of a 5-point scale (1, never, to 5, always).

2.2.1.2.2. Measures. DLC intensity 2 consisted of two measures originally proposed by Hartanto and Yang (2016), i.e., the index of single-language context (SLC) bilingualism and the index of intrasentential code-switching. The index of SLC bilingualism reflects the extent to which a bilingual uses only one language as compared to

[1] Prior to the main study, we carried out a pilot procedure in order to examine the theoretical validity of the questionnaire. It was administered to 30 native-speakers of Polish who declared everyday use of English. First, participants were asked to fill in the paper questionnaire; then, the research assistant conducted one-to-one interviews with every participant separately in order to assess the comprehensibility of the instructions and items. Afterwards, the data collected during the interviews were compared with the data derived from the paper questionnaire. Based on the outcomes of the pilot, the instructions and the items were modified to better elicit the information about the patterns of language use.

[2] In order to test the reliability of language mixing, its internal consistency was assessed using Cronbach's α. The outcomes showed that the responses to statement 1 systematically decreased the internal consistency of the measure, whereas the responses to statement 3 were the most consistent across the contexts. Therefore, the data related to statement 1 were excluded and the data related to statement 3 had double weight.

the overall use of other languages; it was computed following the formula presented in Appendix B.2. of Hartanto and Yang's paper. A lower score of this measure indicates a greater co-occurrence of languages during a day. The index of intrasentential code-switching reflects the frequency of mixing languages within single sentences and corresponds to the language mixing from our questionnaire; it was computed following the formula presented in Appendix A.2. of Hartanto and Yang's paper. A lower score of this measure indicates less frequent mixing of languages within the same utterance.

DLC intensity 2 was computed by summing the standardized values of the index of SLC bilingualism and the index of intrasentential code-switching for each participant separately. Since more intense DLC experience is related to smaller values of these indices, they were reversed before the summation in order to correspond to DLC intensity 1.

2.2.2. Measurement of response inhibition

Response inhibition was measured using four classic tasks: the antisaccade task, the go/no-go task, the Stroop task and the stop-signal task. Each task was preceded by a detailed written instruction and training trials that provided participants with feedback on accuracy. The instructions equally stressed the speed and accuracy of responses. Stimuli did not overlap across the tasks in order to reduce the effects of associative learning on performance. Fig. 1 presents an overview of the response-inhibition tasks.

2.2.2.1. Antisaccade task. The task required participants to indicate the direction of a small arrow. Each trial began with the presentation of a central fixation cross for 500 ms, followed by the presentation of a black square on the left or the right side of the screen that flashed twice for 30 ms with a 16 ms interval of a blank screen in-between. Subsequently, the arrow (pointing either left, up, or right) was presented briefly on the opposite side of the screen for 150 ms, immediately followed by a mask which was presented until a response was made or time ran out (1500 ms).

The fixation cross was 2 mm (0.18°) in width. The arrow, the square and the mask were 3 mm (0.26°) in width. All stimuli in this task were black against a grey background. The size and eccentricity of the arrow were appropriately adjusted in order to make it impossible to identify the direction of the arrow when the participant's gaze fixed in the center of the screen. Participants were instructed to indicate the direction of the arrow (i.e., left, up, or right) by pressing the corresponding arrow key (i.e., "←", "↑", "→", respectively).

Participants were instructed to inhibit a reflexive saccade toward the square and instead make a voluntary saccade to the opposite side in order to identify the direction of the briefly appearing arrow. Before data collection, they received one practice block of 10 trials to ensure that they understood the task. After the practice runs, they completed one experimental block of 50 trials with short breaks in-between. Error rate (antisaccade ERR) was taken as a measure of response inhibition.

2.2.2.2. Stroop task. The task required participants to indicate the color of words while ignoring their meaning. Each trial began with the presentation of a white central fixation cross for 200 ms; this was followed by the presentation of a stimulus in the center of the screen until a response was made or time ran out (1800 ms). The stimuli were four Polish words: blue ("niebieski"), green ("zielony"), red ("czerwony") and yellow ("żółty") displayed in blue, green, red or yellow, respectively. The length of words on the display was 35 mm (3.0°). The fixation cross was 2 mm (0.18°) in width. The stimuli were presented against a black background. In congruent trials, the color of the ink corresponded to the meaning of the word (e.g., the word "red" printed in red). In incongruent trials, the color of the ink did not correspond to the meaning of the word (e.g., the word "red" printed in blue).

In this task, participants were instructed to indicate the color of the ink (i.e., blue, green, red, or yellow) by pressing the corresponding key (i.e., "Z", "X", "N", or "M", respectively) while ignoring the meaning of the word. The stimulus-response mapping was presented at the bottom of the screen during every trial (i.e., names of the colors in the order corresponding to the order of the keys on the keyboard).

Before data collection, participants received one practice block of 14 congruent trials and 10 incongruent trials to ensure that they understood the task. After the practice runs, they completed one experimental block of 160 congruent and 108 incongruent trials, presented in a random order (268 trials in total). Error rate and reaction time in the incongruent condition minus the congruent condition (i.e., the typical Stroop effects) were measures of response inhibition (Stroop ERR and Stroop RT, respectively).

2.2.2.3. Go/no-go task. The task required participants to react to one type of stimulus and inhibit a response to the other type. Each trial began with the presentation of a white central fixation cross for 500 ms, immediately followed by the presentation of a color square until a response was made or time ran out (500 ms). The color of the square indicated the type of trial: red indicated the no-go trial, whereas green and orange indicated frequent-go and infrequent-go trials, respectively. The stimuli were shown against a black background. The fixation cross was 2 mm (0.18°) in width, whereas the square was 3 mm (0.26°) in width. Participants were instructed to react to the go (i.e., green or orange square) by pressing the enter key using the right thumb.

Before data collection, participants received one practice block of 15 frequent-go, three infrequent-go and three no-go trials. After the practice runs they completed five experimental blocks of 150 frequent-go, 25 infrequent-go and 25 no-go trials each, presented in a random

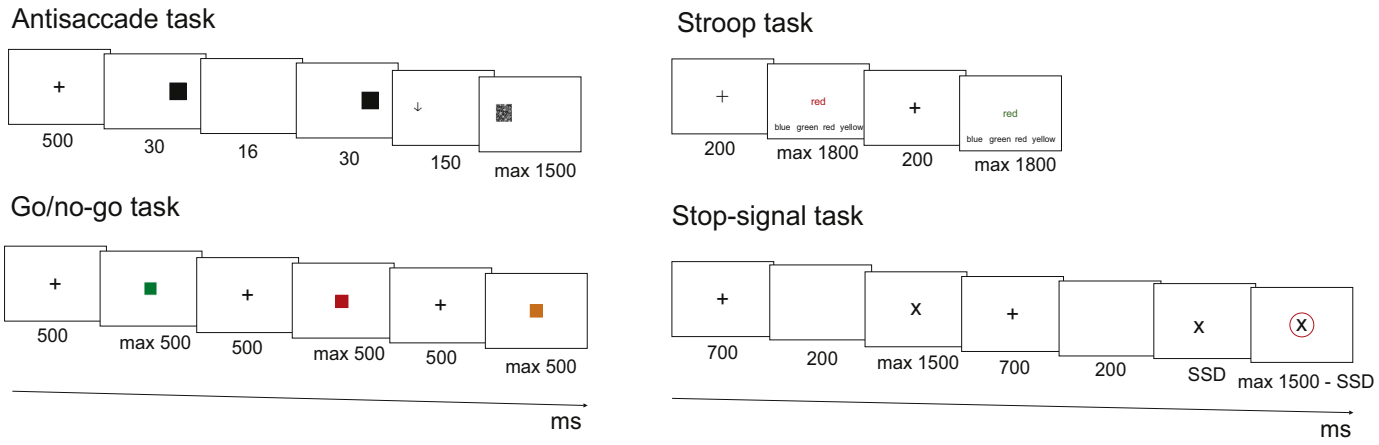

Fig. 1. Exemplar trials for the response-inhibition tasks. Note. SSD = Stop-Signal Delay.

order (500 trials in total). Error rate on the no-go trials (no-go ERR) was taken as a measure of response inhibition.

2.2.2.4. Stop-signal task. The task required participants to perform a primary binary-choice response task. Each trial began with the presentation of a central fixation cross for 700 ms, immediately followed by the presentation of a black screen for 200 ms. Afterwards, a go stimulus was presented in the center of the screen until a response was made or timeout was reached (1500 ms). The go stimulus was one of two letters (either "X", or "O") with 50% probability for each of them. The stimuli were shown in white against a black background. The length of the letter in the display was 3 mm (0.26°). The fixation cross was 2 mm (0.18°) in width. Participants were instructed to indicate the correct letter (i.e., "X" or "O") by pressing the corresponding Ctrl key (i.e., left or right, respectively) using their index fingers.

In a random sample of 25% of trials, a red circle surrounded the presented letter and served as the stop signal in the task. The red circle prompted participants to inhibit their responses to the primary go task, regardless of the letter type. The interval between the presentation of the go stimulus and the stop signal (i.e., the stop-signal delay, SSD) was varied trial by trial using a tracking method: the interval either increased or decreased by 50 ms (from 100 to 400 ms) for the next stop signal trial, depending on whether participants either successfully or unsuccessfully inhibited their response to the go stimulus, respectively. Thus, there were seven possible SSDs: 100, 150, 200, 250, 300, 350, and 400 ms. After a successful inhibition, the interstimulus interval became longer; after an unsuccessful inhibition, it became shorter. The initial value of the SSD was set to 150 ms. The tracking method aimed to converge on an SSD at which participants successfully inhibited responses to approximately 50% of the stop-signal trials.

Before data collection, participants received one practice block of 20 go trials and five stop signals to ensure that they understood the task. After the practice runs, they completed four experimental blocks, each consisting of 50 trials with short breaks between the blocks. During the break, the accuracy feedback for stop-trials and mean reaction time were presented centrally on the screen.

Stop-signal reaction time (SSRT) was taken as a measure of response inhibition. It was calculated following the procedure of Logan (1994). RT from go stimuli responses on which no stop signal occurred were collapsed into a single distribution and rank ordered. The *n*th RT was selected, where n was obtained by multiplying the number of no-signal RTs in the distribution (150) by the probability of responding (e.g., 0.5 if the inhibition rate in the task was 50%) for each participant separately. SSRT was calculated by subtracting the average stop-signal delay (SSD) from the *n*th RT, following the horse race model (for more detail, see Logan & Cowan, 1984; Verbruggen & Logan, 2008).

2.3. General procedure

The participants were tested in groups of up to eight during a session of approximately 2–2.5 h (including breaks between blocks of tasks and a longer break in the middle of a session). After informed consent was obtained, the participants performed the battery of tasks and questionnaires administered in fixed order: antisaccade task, written fluency task, LexTALE task, Stroop task, Patterns of Language Use Questionnaire, go/no-go task, fluid intelligence test, Code-switching and Interactional Contexts Questionnaire, stop-signal task, socio-demographic background questionnaire, language background questionnaire. The data concerning the written fluency task were beyond the scope of this study.

2.4. Data preparation and analyses

2.4.1. Measures of response inhibition

For the antisaccade task, the go/no-go task, and the Stroop task, all trials with timeouts or extremely short reaction times (RTs below 100 ms in the go/no-go task and below 250 ms in the other tasks) were removed.[3] The Stroop RT was computed from accurate trials only that were within 3 standard deviations of the mean for a given participant. Due to a skewed data distribution, we applied the logarithm transformation to RT-based measures (i.e., Stroop RT and SSRT) and the arcsine square root transformation to ERR-based measures (i.e., no-go ERR, antisaccade ERR and Stroop ERR).

2.4.2. Factor analysis

The data were analyzed using R (R Core Team, 2019). First, the psychometric properties of the DLC intensity measures (language entropy, language mixing, index of SLC bilingualism, index of intrasentential code-switching) and the response-inhibition measures (antisaccade ERR, no-go ERR, Stroop ERR, Stroop RT, SSRT) were assessed. The reliability of the language mixing measure was tested using the Cronbach's α (the reliabilities could not be assessed for language entropy, index of SLC bilingualism, index of intrasentential code-switching, as they consisted of single values). The reliabilities of the response inhibition measures were assessed using the split-half correlations. The correlations were computed between odd and even items within the task conditions that included conflict and adjusted using the Spearman–Brown prophecy formula. Before correlations, all measures were centered and scaled in order to ensure a common measurement scale. In addition, multivariate normality for response-inhibition measures was evaluated using the Mahalanobis distance, and participants classified as multivariate outliers were excluded. Next, Exploratory Factor Analysis (EFA) was used to determine whether the response-inhibition measures enabled identification of a common factor. The Structural Equation Modeling (SEM) was used to test the relationship between DLC intensity and response inhibition. The SEM models were fitted using the lavaan package (Rosseel, 2012). Their fit was evaluated using the following indices: chi-square statistic (χ^2; value should be non-significant), Bentler's comparative fit index (CFI > 0.95), the root mean square error of approximation (RMSEA <0.06), and the standardized root-mean-square residual (SRMR <0.08) (for discussion see Kline, 2016). The data and the R scripts to generate the models are available at https://osf.io/5x9sc/.

3. Results

3.1. Descriptive statistics, reliabilities and correlations for DLC intensity measures

Table 4 presents descriptive statistics for the DLC intensity measures. On average, the participants were likely to use both languages in the same context (as indicated by relatively high language entropy and relatively low index of SLC bilingualism) and they moderately mixed languages within single utterances (based on the language mixing and the index of intrasentential code-switching). Language mixing demonstrated satisfactory reliability (ranging from α = 0.82 for the home context up to α = 0.87 for the free-time context).

Table 5 presents the correlation coefficients between the measures of DLC intensity. Overall, the correlations were significant and their directions were in line with the predictions. The corresponding measures across the questionnaires (i.e., language entropy and SLC bilingualism; language mixing and intrasentential code-switching) demonstrated strong correlations. Altogether, the correlation matrix indicated that these measures measured one and the same construct.

Fig. 2 presents the correlation between language entropy and

[3] Since the go/no-go task was highly speeded in this study, we used different cut-off for removing anticipatory responses in this task than in the Stroop task. Based on a visual screening of the data, we decided to use 100 ms as a cut-off point.

Table 4
Descriptive statistics of the DLC intensity measures (N = 195).

Questionnaire	Measure	*Mean*	*SD*	Min	Max	Skew	Kurtosis
Patterns of language use	Language entropy	0.55	0.20	0.00	1.10	0.11	0.09
	Language mixing	4.80	1.94	0.00	8.56	−0.04	−0.91
Code-switching and interactional contexts	SLC bilingualism	0.47	0.21	0.03	1.00	0.22	−0.64
	Intrasentential code-switching	4.74	2.11	0.00	9.00	−0.11	−0.96

Note. *SD*, standard deviation.

Table 5
Correlations for the DLC intensity measures (N = 195).

Measure	Language entropy	Language mixing	SLC bilingualism	Intra code-switch	DLC intensity 1
Language mixing	0.19 [0.04, 0.34]				
SLC bilingualism	−0.90 [−0.93, −0.88]	−0.19 [−0.34, −0.03]			
Intra code-switch	0.38 [0.25, 0.50]	0.65 [0.54, 0.74]	−0.38 [−0.50, −0.25]		
DLC 1	0.64 [0.55, 0.71]	−0.64 [−0.71, −0.55]	−0.56 [−0.65, −0.46]	−0.21 [−0.35, −0.06]	
DLC 2	0.47 [0.37, 0.56]	−0.41 [−0.52, −0.29]	−0.56 [−0.63, −0.47]	−0.56 [−0.64, −0.47]	0.63 [0.62, 0.76]

Notes. *DLC 1*, a measure of DLC intensity based on language entropy and language mixing; *DLC 2*, a measure of DLC intensity based on SLC bilingualism and intrasentential code-switching; *intra code-switch*, intrasentential code-switching; all measures were standardized; upper and lower CI from a bootstrapping procedure with 10,000 random samples are given in brackets; for all correlations, $p \leq 0.02$.

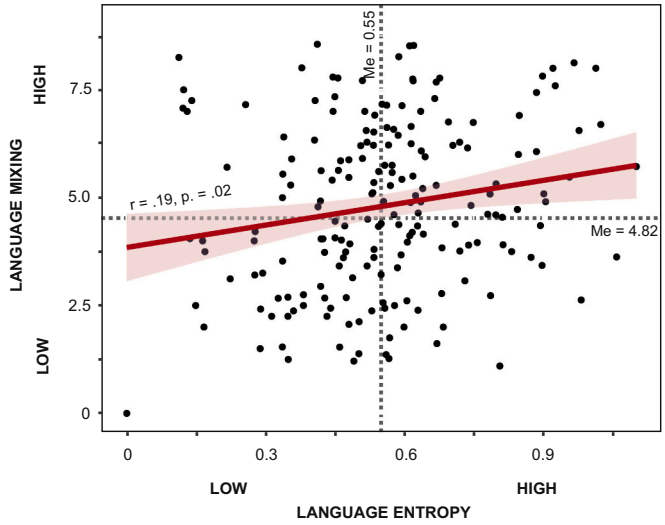

Fig. 2. Participants' patterns of language use. Correlation between language entropy (raw data; x axis) and language mixing (raw data; y axis); solid red line indicates least squares fit with 95% CI (red belt); grey dotted lines indicate median (Me) values for the indices. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

language mixing. As can be seen, the participants varied in their everyday patterns of language use and represented a wide spectrum of dependencies between the co-occurrence of languages and the frequency of language mixing.

3.2. Descriptive statistics, reliabilities and correlations for response-inhibition measures

Table 6 presents descriptive statistics and reliabilities for the response-inhibition measures. For the Stroop task, go/no-go task and stop-signal task, the split-half reliability estimates indicated excellent reliability. For the antisaccade task, the estimates were lower but still indicated very good reliability.

Table 7 presents the correlations between the response-inhibition measures. Before the correlation analysis, we excluded the data of nine participants who incorrectly performed the stop-signal task and seven participants who were classified as multivariate outliers based on the Mahalanobis distance, thus leaving 179 participants in the analyses. Overall, the correlations were either low or non-significant, even for measures derived from the same task (i.e., Stroop RT and Stroop ERR). Moreover, although SSRT correlated with antisaccade ERR, the direction of this correlation was opposite to what was expected.

Altogether, the correlation matrix suggested weak and null associations between the response-inhibition measures. Consistently with the correlational analysis, the common factor in the EFA (see Table 8) explained only 11% of the observed variance. SSRT did not load onto the factor, which suggests that the stop-signal task did not measure the same construct as did the other tasks. Therefore, SSRT was excluded from the further analyses.

3.3. SEM results

In this analysis, nine participants who incorrectly performed the stop-signal task were re-included as this task did not enter SEM. Due to inclusion of additional participants, the multivariate normality of the response-inhibition data was re-evaluated. Based on the Mahalanobis distance, seven participants were excluded, leaving 188 participants in the SEM analysis. Table 9 presents the factor loadings for the response-inhibition measures and the correlations for the SEM dataset.

To test the relationship between DLC intensity and response inhibition, we fitted the model in which the DLC intensity factor was considered as the predictor of the response-inhibition factor (Model 1). The model included the score on the fluid intelligence test and the SES factor as additional predictors of response inhibition (covariates) to control for individual differences in the participants' background characteristics. The DLC intensity factor consisted of two measures, namely DLC intensity 1 and DLC intensity 2, whereas the response-inhibition factor consisted of the four response-inhibition measures (i.e., antisaccade ERR, Stroop ERR, Stroop RT and no-go ERR). Parental education, social status and income represented the SES factor. A series of additional models were fitted to see how the exclusion of covariates affects the relationship between DLC intensity and response inhibition. Model 2 excluded intelligence; Model 3 did not include SES, whereas Model 4 excluded both. The four models were estimated using the maximum likelihood method (Schumacker & Lomax, 2004). The predictors were uncorrelated in the models, reflecting the respective null correlations among them. The variance of the factors was fixed to 1.0 and the factor loadings of DLC intensity 1 and DLC intensity 2 were constrained to be equal in order to ensure that they equally loaded onto the DLC intensity factor. The residual errors of Stroop RT and Stroop ERR were correlated, as these measures were derived from the same

Table 6
Descriptive statistics and reliabilities of the response-inhibition measures.

	Measure	*n*	*Mean*	*SD*	Min	Max	Skew	Kurtosis	Reliability
Antisaccade task	Antisaccade ERR	195	0.21	0.15	0.00	0.68	0.61	0.07	0.86
Stroop task	Stroop ERR	195	0.07	0.13	−0.05	0.93	5.05	29.23	0.96[a]
	Stroop RT	195	175.43	84.86	−43.50	434.93	0.24	0.04	0.92[a]
Go/no-go task	No-go ERR	195	0.35	0.16	0.00	0.77	0.22	−0.73	0.93[b]
Stop-signal task	SSRT	186	249.06	56.52	115.61	435.54	0.68	0.53	0.99[c]

Notes. *SD*, standard deviation.
[a] Split-half reliability of ERR/RT in the incongruent condition.
[b] Split-half reliability of ERR in the no-go condition.
[c] Split-half reliability of RTs in the go condition.

task.

Table 10 describes the fit of the models to the data. All models yielded a very good fit. Fig. 3 presents Model 1. The measures of DLC intensity, SES and inhibition significantly loaded on the respective factors. The loadings on DLC intensity and SES were satisfactory. However, the loadings on response inhibition were relatively low, which indicates that the response-inhibition measures weakly contributed to the common factor. Intelligence and SES influenced response inhibition, indicating that individuals with higher intelligence and higher socio-economic status demonstrate better response-inhibition skills. However, contrary to the main prediction of this study, no significant relationship was observed between DLC intensity and response inhibition.

In Models 2 and 3, the effects of covariates (SES and intelligence, respectively) on response-inhibition factor were still significant and none of the additional models showed a significant relationship between DLC intensity and response inhibition. Taken together, the outcomes of the SEM analyses consistently indicate that the differences in the intensity of DLC experience did not influence response inhibition.

3.4. Exploratory analyses

The SEM analysis did not show evidence for the relationship between DLC intensity and response inhibition. However, because the response-inhibition factor represented limited common variance across the measures, it could be argued that although the single tasks measured some aspects of response inhibition related to the language use of bilinguals, the SEM model might have filtered this variance out. In order to test this possibility, the ACH's prediction was tested against each of the response-inhibition measures separately. While such individual variables are subject to high measurement error (see Section 1.2), this test could potentially guide further research into the relationship between bilingualism and response inhibition.

3.4.1. Frequentist linear regression for the single measures of response inhibition

For each response-inhibition measure (i.e., antisaccade ERR, Stroop ERR, Stroop RT, no-go ERR), we fitted a linear regression model corresponding to Model 1 tested in the SEM analysis (i.e., the DLC intensity factor, the SES factor and the score on the fluid intelligence test as predictors). The analysis was performed on the same dataset as used in the SEM analysis (N = 188) and a total of four models were tested. The model was statistically significant for antisaccade ERR, F (3,184) = 4.16, p = 0.007, R^2 = 0.06, but not for the other measures ($Fs \leq 2.24$) (see also Table 11).

Intelligence and SES predicted antisaccade ERR, suggesting that more effective inhibition of reflexive saccades is related to higher intelligence and/or better socio-economic status. Crucially, however, there was no effect of DLC intensity on antisaccade ERR. As such, the outcomes of the linear regression analysis agree with those of the SEM analysis and suggest no relationship between the intensity of DLC experience in bilinguals and their response-inhibition skills.

Table 8
Exploratory factor analysis for the measures of response inhibition (N = 179).

Measure	Factor loading	Uniqueness
Antisaccade ERR	0.15	0.98
Stroop ERR	0.28	0.92
Stroop RT	0.57	0.68
No-go ERR	0.36	0.87
SSRT	–	0.99

3.4.2. Bayesian linear regression for the single measures of response inhibition

In order to assess the evidence against the effect of DLC intensity in the response-inhibition measures, the Bayesian-estimation approach was employed. For each response-inhibition measure, the evidence in favor of the absence of the DLC intensity effect (i.e., in favor of the null hypothesis) was assessed using Bayes factors as implemented in the brms package (version 2.11.1; Bürkner, 2018). The Bayes factors were computed for the contrast between the two models. The first model resembled the model fitted using the frequentist regression and SEM, whereas the second model had the DLC intensity factor excluded. Table 12 display the 95% credible interval for the DLC intensity effect in the first model and the Bayes factor in favor of the second model (i.e., model without the DLC intensity factor). As can be seen, zero was included in all credible intervals which suggested that there was no effect of DLC intensity in the data. The Bayes factors were stable (i.e., did not change when re-computed) and suggested substantial evidence (i.e., BF between 3 and 10; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011) against the DLC intensity effect for each of the response-

Table 7
Correlations for the response-inhibition measures (N = 179).

Measure	Antisaccade ERR	Stroop ERR	Stroop RT	No-go ERR
Stroop ERR	0.05 [−0.10, 0.20]	–		
Stroop RT	0.05 [−0.09, 0.18]	0.16 [−0.06, 0.39]	–	
No-go ERR	0.08 [−0.07, 0.23]	0.10 [−0.05, 0.24]	**0.21** [0.07, 0.34]	–
SSRT	**−0.22** [−0.36, −0.07]	0.01 [−0.12, 0.15]	−0.08 [−0.22, 0.06]	0.07 [−0.07, 0.21]

Notes. All measures were standardized; upper and lower CI from a bootstrapping procedure with 10,000 random samples are given in brackets; bold font indicates $p < 0.05$.

Table 9
Factor loadings and correlations after exclusion of SSRT (N = 188).

Measure	Loading (uniqueness)	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Antisaccade ERR [1]	**0.16 (0.98)**	–	–	–	–	–	–	–	–
Stroop ERR [2]	**0.46 (0.79)**	0.08 [−0.07, 0.22]	–	–	–	–	–	–	–
Stroop RT[3]	**0.64 (0.59)**	0.08 [−0.05, 0.22]	**0.30 [0.17, 0.42]**	–	–	–	–	–	–
No-go ERR [4]	**0.34 (0.86)**	0.10 [−0.05, 0.24]	**0.14 [0.02, 0.26]**	**0.22 [0.08, 0.35]**	–	–	–	–	–
Ladder [5]	–	−0.09 [−0.24, 0.05]	−0.15 [−0.30, 0.01]	−0.05 [−0.20, 0.11]	−0.10 [−0.25, 0.04]	–	–	–	–
Income [6]	–	−0.12 [−0.26, 0.02]	−0.02 [−0.17, 0.13]	−0.04 [−0.18, 0.10]	−0.10 [−0.23, 0.03]	**0.32 [0.17, 0.46]**	–	–	–
Parental education [7]	–	**−0.22 [−0.35, −0.08]**	−0.13 [−0.27, 0.01]	−0.13 [−0.27, 0.02]	−0.01 [−0.15, 0.14]	**0.25 [0.13, 0.36]**	**0.20 [0.06, 0.33]**	–	–
DLC 1 [8]	–	0.06 [−0.07, 0.19]	−0.02 [−0.15, 0.11]	0.03 [−0.11, 0.16]	0.02 [−0.15, 0.19]	0.03 [−0.10, 0.16]	−0.03 [−0.17, 0.10]	−0.02 [−0.16, 0.13]	–
DLC 2 [9]	–	0.03 [−0.11, 0.17]	−0.06 [−0.17, 0.06]	−0.09 [−0.22, 0.03]	−0.01 [−0.15, 0.14]	0.11 [−0.02, 0.24]	0.01 [−0.12, 0.14]	−0.02 [−0.17, 0.13]	**0.71 [0.63, 0.77]**

Notes. *DLC 1*, a measure of DLC intensity based on language entropy and language mixing; *DLC 2*, a measure of DLC intensity based on SLC bilingualism and intrasentential code-switching; all measures were standardized; upper and lower CI from a bootstrapping procedure with 10,000 random samples are given in brackets; bold font indicates $p < 0.05$.

Table 10
Structural Equation Modeling statistics and goodness-of-fit for models predicting the response-inhibition factor (N = 188).

	χ^2	*df*	*p*	CFI	RMSEA [90% CI]	SRMR	DLC intensity path to response inhibition
Model 1	32.27	33	0.50	1.00	0.00 [0.00, 0.05]	0.05	−0.03 [−0.29, 0.32]
Model 2	23.91	25	0.52	1.00	0.00 [0.00, 0.06]	0.05	−0.03 [−0.32, 0.25]
Model 3	12.77	13	0.47	1.00	0.00 [0.00, 0.07]	0.04	−0.04 [−0.29, 0.21]
Model 4	5.90	8	0.66	1.00	0.00 [0.00, 0.07]	0.03	−0.07 [−0.31, 0.18]

Notes. *Model 1* = intelligence, SES and DLC intensity as predictors of response inhibition; *Model 2* = SES and DLC intensity as predictors; *Model 3* = intelligence and DLC intensity as predictors; *Model 4* = DLC intensity as a single predictor; upper and lower CI from a bootstrapping procedure with 10,000 random samples are given in brackets; note that the relationships between DLC intensity and response inhibition were non-significant in all models.

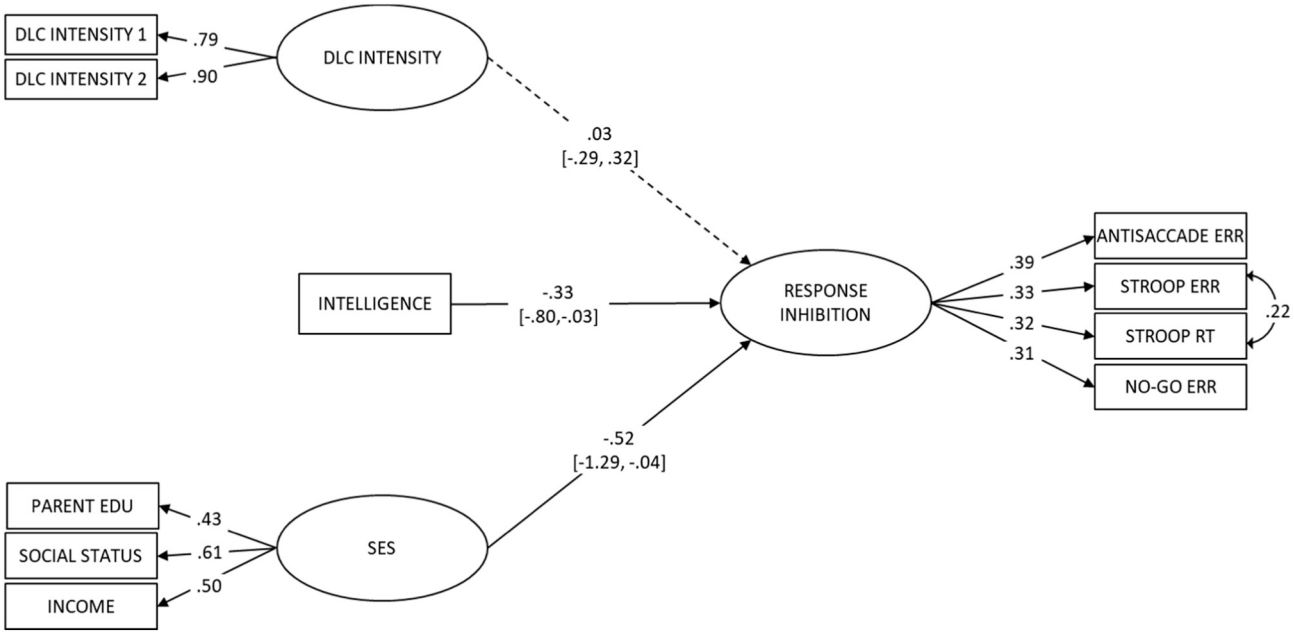

Fig. 3. Structural Equation Model 1 with intelligence, SES and DLC intensity as predictors of response inhibition. *Notes*. *DLC intensity 1*, a measure of DLC intensity based on language entropy and language mixing; *DLC intensity 2*, a measure of DLC intensity based on SLC bilingualism and intrasentential code-switching; *Parent edu*, parental education; ovals represent latent variables (factors); boxes represent manifest variables (single measures); number next to the rounded line is the covariance coefficient for residual errors; numbers next to the short arrows are the standardized factor loadings; numbers next to the long arrows are the standardized path coefficients (regression weights); solid lines indicate $p < 0.05$; the dotted line indicates $p = 0.81$.

Table 11
Estimates for the classic regression model for antisaccade ERR.

	B	*SE*	*t*	*p*
Intercept	−0.02	0.07	−0.22	0.83
DLC intensity	0.06	0.07	0.88	0.38
Socio-economic status	−0.22	0.10	−2.35	0.02
Intelligence	−0.18	0.07	−2.44	0.02

Table 12
Outcomes of the Bayesian analyses regarding the absence of the DLC intensity effect.

Measure	95% CI for DLC intensity effect	BF01
Antisaccade ERR	[−0.08, 0.21]	3.78
Stroop ERR	[−0.11, 0.07]	7.99
Stroop RT	[−0.21, 0.06]	3.25
No-go ERR	[−0.14, 0.15]	5.50

Note. *95% CI for DLC intensity effect* = 95% credible interval for the DLC intensity effect in model including the DLC intensity factor; *BF01* = evidence for the model without the DLC intensity factor over the model including the DLC intensity factor.

inhibition measures tested in this study.

The results of the Bayesian analysis complements the outcomes of the SEM and the single-task linear regressions and indicate that the DLC intensity did not influence the participants' performance in the four classic response-inhibition tasks. Taken together, the presented analyses speak against the ACH's prediction: the DLC experience does not seem to confer an advantage in response inhibition in bilingual speakers.

4. Discussion

The goal of the current study was to test the Adaptive Control Hypothesis (ACH), which posits that bilingualism enhances only these cognitive control processes that are actively used by bilinguals to control their concurrent use of two languages (Green & Abutalebi, 2013). To this aim, we conducted a large-scale correlational study in a group of Polish-English bilinguals living in Poland and using two languages on an everyday basis. We focused on the process of response inhibition, which according to the ACH should be intensively trained and enhanced in bilinguals who use two languages in the same context but do not mix these languages within the same utterance (i.e., DLC bilinguals).

Based on the ACH, we defined the intensity of DLC experience in bilinguals as the interplay between the co-occurrence of languages (i.e., the extent to which two languages are used within the same context) and the frequency of language mixing (i.e., the extent to which two languages are mixed within the same utterance). We also proposed a new way of estimating the co-occurrence of languages by measuring the probability distribution of all languages used within and across contexts during a day (language entropy; for more detail see Section 2.2.1.1; see also Gullifer & Titone, 2019). The response-inhibition construct was measured by tasks which have been shown to form the response-inhibition factor in previous work (Chuderski et al., 2012; Friedman & Miyake, 2004; Miyake et al., 2000). By testing a more nuanced version of the bilingual advantage hypothesis (the ACH) in a relatively large sample of participants and by employing the latent-variable approach, we addressed some methodological shortcomings of previous studies that tested the bilingual advantage hypothesis. Based on the ACH predictions, we hypothesized that the higher the intensity of DLC experience for a given bilingual, the better their response inhibition.

4.1. Convergent validity of theoretical constructs

The latent-variable approach allowed reliable and valid measurement of theoretical constructs, i.e., DLC intensity and response inhibition (as recommended by Friedman, 2016; Gade, 2015; Paap, 2014). In order to measure the intensity of DLC experience, we used the questionnaire developed in this study and also the questionnaire devised by Hartanto and Yang (2016). The questionnaire-derived measures strongly correlated with each other and ascertained the credible measurement of DLC intensity. Crucially, the measures showed that the participants considerably varied in their everyday patterns of language use.

In order to measure response inhibition, we employed a battery of four classic tasks: the antisaccade task, the go/no-go task, the Stroop task, and the stop-signal task, that provided five different measures of response inhibition. Although all measures revealed satisfactory reliabilities, their inter-correlations were weak or non-significant even when two measures derived from the same task were taken into account (i.e., Stroop effect measured in RT and ERR). In addition, the performance in the stop-signal task (i.e., as indexed by SSRT) negatively correlated with error rate in the antisaccade task (instead of expected positive correlation) and did not load onto the common factor in the EFA. The SEM analysis showed that the response-inhibition measures weakly contributed to the common factor. As such, although the model fitted the data well, the response-inhibition factor represented a limited amount of common variance among the response-inhibition tasks applied.

Weak associations between the response-inhibition tasks cannot be easily explained by the idiosyncratic properties of our study. First, we used the tasks that formed the latent variables in previous work (Chuderski et al., 2012; Friedman & Miyake, 2004; Miyake et al., 2000). Second, all the tasks had reliabilities that were as good as or even better than those found in previous research (for a comparison of the reliability estimates in previous research, see Table 8 in Rey-Mermet et al., 2018, p. 15); as such, the tasks provided reliable measures of the participants' performance. Third, although we tested young adults who are often argued to be susceptible to ceiling effects in cognitive performance (Bialystok, 2017; Bialystok, Abutalebi, Bak, Burke, & Kroll, 2016; Bialystok, Martin, & Viswanathan, 2005), there was a considerable variability in the performance of response-inhibition tasks (for similar findings see Samuel et al., 2018). Therefore, the limited common variance as represented by the response-inhibition factor cannot be attributed to the restricted variance in the individual response-inhibition measures. What is more, the limited common variance between the response-inhibition measures observed in this study is in line with several recent reports that tested the psychometric properties of the response-inhibition construct (Paap, Anders-Jefferson, Zimiga, Mason, & Mikulinsky, 2020; Rey-Mermet et al., 2018; Rouder, Kumar, & Haaf, 2019). Rey-Mermet et al. (2018) used SEM to test the impact of age on the efficiency of inhibitory control. Similarly to our study, they also administered the antisaccade, stop-signal, and Stroop task, and observed low or non-significant correlations between the response-inhibition measures.

To conclude, reliable measures of the patterns of language use allowed a robust measurement of DLC intensity to be obtained. Importantly, to the best of our knowledge this is the first study to show that the bilinguals' patterns of language use (also called interactional context) can be represented at the latent level. At the same time, however, the measures of response inhibition, although highly reliable, did not form a robust latent variable, which suggests that the classic response-inhibition tasks primarily measure task-specific processes rather than a unitary construct of response inhibition. Notably, however, despite the limitation of common variance in response inhibition, the factor captured an underlying ability to some extent, as suggested by its significant relationships with intelligence and social status.

4.2. Does the intensity of DLC experience predict the efficiency of response inhibition?

The SEM analysis aimed to determine whether the efficiency of response inhibition can be predicted by the intensity of the DLC experience while controlling for variables that could confound this relationship, i.e., intelligence and socio-economic status (SES). Significant effects of SES and intelligence on response inhibition are in line with previous studies showing better efficiency of cognitive functions in individuals with higher SES (Arán-Filippetti & de Minzi, 2012; Ursache & Noble, 2016) and intelligence (Friedman et al., 2006; see also Chuderski et al., 2012). Crucially, however, DLC intensity did not predict response inhibition, indicating that the efficiency of response inhibition was not related to the intensity of DLC experience in bilinguals. As such, we did not find support for the prediction of the ACH using SEM. However, since the response-inhibition factor represented limited common variance across the measures, it could be argued that the tasks measured some aspects of response inhibition related to the language use of bilinguals, but the SEM model did not account for this variance. To address this issue, we tested the ACH's prediction in each of the response-inhibition measures individually using frequentist and Bayesian approaches. Consistently with the SEM outcomes, the follow-up analyses also showed that the DLC intensity did not predict performance in the response-inhibition tasks.

Overall, the outcomes of all analyses consistently speak against the prediction of ACH and suggest that there is no relationship between the intensity of DLC experience and the efficiency of response inhibition at the behvioral level. Importantly, the null relationship between DLC intensity and response inhibition cannot be explained by the insufficient reliability of the data in this study. Relatively high reliabilities of the measures ascertain that the absence of the expected effects was not a measurement error in this study. Therefore, the most straightforward interpretation of our results is that, in contrast to the prediction of the ACH, bilinguals who operate in DLC either do not engage response inhibition to control language production or use it to the same extent as other bilinguals who either mix two languages freely or use only one language in a given context. It should be kept in mind, however, that the current investigation was limited to response-inhibition and therefore it does not exclude the possibility that a variety of bilinguals' patterns of language use affect other cognitive processes considered in the ACH (e.g., switching in Hartanto & Yang, 2020).

5. Conclusions

In this study, we tested one of the predictions of the Adaptive Control Hypothesis, according to which the use of different languages in the same situation without mixing them in single utterances (called dual-language context) engages and consequently trains response inhibition in bilingual speakers. We attempted to circumvent several pitfalls of previous research on the cognitive benefits of bilingualism by testing a relatively large sample of participants and by employing a more reliable and valid measurement of constructs (i.e., the latent variable approach accompanied by Bayesian estimation). Although the study provided highly reliable measures of bilingualism and response inhibition, the results do not support the prediction of the ACH: the intensity of using different languages in the same context without mixing them in single utterances was unrelated to the efficiency of response inhibition, regardless of whether inhibition was estimated using the latent variable approach or single measures. As such, the results suggest that bilinguals who operate in a dual-language context either do not engage response inhibition to control language production or engage it to the same extent as other bilinguals. Therefore, we conclude that the ACH probably does not account for the discrepant results of studies testing the relationship between bilingualism and cognitive control efficiency, at least with respect to response inhibition.

Importantly, the study also highlights the problem of measuring response inhibition at the behavioral level, as we observed only a moderate commonality among the response inhibition tasks. No evidence for a robust response-inhibition construct adds to the growing skepticism on this issue in the literature (Haaf & Rouder, 2019; Karr et al., 2018; Paap et al., 2020; Rey-Mermet et al., 2018; Rouder & Haaf, 2019). If response inhibition indeed forms a unitary construct, future research will need to reevaluate the methods of its measurement. Finally, the present study draws attention to the problem of reliable and valid measurement of the relationship between bilingualism and cognitive control. Because the available evidence for the cognitive effects of bilingualism mostly comes from studies that adopted a single-task approach (i.e., one construct – one measure) and did not report the psychometric properties of the measures, it is not clear whether the findings reflect individual differences in the assumed underlying constructs or merely task-specific effects. The absence of valid and reliable behavioral methods of measuring individual differences in cognitive control poses a serious challenge for testing any hypotheses related to differences in these abilities.

CRediT authorship contribution statement

Patrycja Kałamała: Conceptualization, Investigation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition, Formal analysis. **Jakub Szewczyk**: Conceptualization, Writing - original draft, Writing - review & editing, Supervision, Formal analysis. **Adam Chuderski**: Conceptualization, Writing - original draft, Writing - review & editing, Supervision, Formal analysis. **Magdalena Senderecka**: Conceptualization, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition. **Zofia Wodniecka**: Conceptualization, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgements

The research was funded by a Ministry of Science and Higher Education grant (0002/DIA/2014/43). During work on the paper, P.K. was supported by a National Science Centre grant (2017/27/N/HS6/01029), Z.W., J.S., and P.K. were supported by a National Science Centre grant (2015/18/E/HS6/00428), and M.S. was supported by a National Science Centre grant (2015/19/B/HS6/00341). The authors gratefully acknowledge the help of Zuzanna Skóra with Bayesian analysis, Joanna Durlik with designing the questionnaire and Michael Timberlake with proofreading.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2020.104373.

References

Antoniou, M. (2019). The advantages of bilingualism debate. *Annual Review of Linguistics, 5*(1), 395–415. https://doi.org/10.1146/annurev-linguistics-011718-011820.

Arán-Filippetti, V., & de Minzi, M. C. R. (2012). A structural analysis of executive functions and socioeconomic status in school-age children: Cognitive factors as effect mediators. *The Journal of Genetic Psychology, 173*(4), 393–416. https://doi.org/10.1080/00221325.2011.602374.

Bak, T. H. (2015). Beyond a simple “yes” and “no”. *Cortex, 73*, 332–333. https://doi.org/10.1016/j.cortex.2015.08.003.

Bak, T. H. (2016a). The impact of bilingualism on cognitive ageing and dementia: Finding a path through a forest of confounding variables. *Linguistic Approaches to Bilingualism, 6*(1–2), 205–226. https://doi.org/10.1075/lab.15002.bak.

Bak, T. H. (2016b). Cooking pasta in La Paz: Bilingualism, bias and the replication crisis. *Linguistic Approaches to Bilingualism, 6*(5), https://doi.org/10.1075/lab.16002.bak.

Bakker, M. (2015). Power problems: N > 138. *Cortex, 73*, 367–368. https://doi.org/10.1016/j.cortex.2015.07.006.

Barbu, C., Orban, S., Gillet, S., & Poncelet, M. (2018). The impact of language switching frequency on attentional and executive functioning in proficient bilingual adults. *Psychologica Belgica, 58*(1), 115–127. https://doi.org/10.5334/pb.392.

von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive

advantages: A test of four hypotheses. *Journal of Experimental Psychology: General, 145*(2), 246–258. https://doi.org/10.1037/xge0000120.
Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language, 95*, 173–189. https://doi.org/10.1016/j.jml.2017.04.002.
Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2019). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000770.
Bialystok, E. (2016). The signal and the noise: Finding the pattern in human behavior. *Linguistic Approaches to Bilingualism, 6*(5), 517–534. https://doi.org/10.1075/lab.15040.bia.
Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin, 143*(3), 233–262. https://doi.org/10.1037/bul0000099.
Bialystok, E., Abutalebi, J., Bak, T. H., Burke, D. M., & Kroll, J. F. (2016). Aging in two languages: Implications for public health. *Ageing Research Reviews, 27*, 56–60. https://doi.org/10.1016/j.arr.2016.03.003.
Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism, 9*, 103–119. https://doi.org/10.1177/13670069050090010701.
Blanco-Elorrieta, E., & Pylkkänen, L. (2018). Ecological validity in bilingualism research and the bilingual advantage. *Trends in Cognitive Sciences, 22*(12), 1117–1126. https://doi.org/10.1016/j.tics.2018.10.001.
Boumeester, M., Michel, M. C., & Fyndanis, V. (2019). Sequential multilingualism and cognitive abilities: Preliminary data on the contribution of language proficiency and use in different modalities. *Behavioral Science, 9*(9), 92. https://doi.org/10.3390/bs9090092.
de Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Science, 9*(3), https://doi.org/10.3390/bs9030033.
Bürkner, P. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal, 10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017.
Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence, 40*(3), 278–295. https://doi.org/10.1016/j.intell.2012.02.010.
Congdon, E., Mumford, J. A., Cohen, J. R., Galvan, A., Canli, T., & Poldrack, R. A. (2012). Measurement and reliability of response inhibition. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00037.
Dick, A. S., Garcia, N. L., Pruden, S. M., Thompson, W. K., Hawes, S. W., Sutherland, M. T., Riedel, M. C., Laird, A. R., & Gonzalez, R. (2019). No evidence for a bilingual executive function advantage in the nationally representative ABCD study. *Nature Human Behaviour, 3*, 692–701. https://doi.org/10.1038/s41562-019-0609-3.
Donnelly, S., Brooks, P. J., & Homer, B. (2015). Examining the bilingual advantage on conflict resolution tasks: A meta-analysis. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*https://mindmodeling.org/cogsci2015/papers/0111/paper0111.pdf.
Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116–124. https://doi.org/10.3758/BF03195503.
Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*(2), 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x.
Friedman, N. P. (2016). Research on individual differences in executive functions: Implications for the bilingual advantage hypothesis. *Linguistic Approaches to Bilingualism, 6*(5), 535–548. https://doi.org/10.1075/lab.15041.fri.
Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General, 133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101.
Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023.
Gade, M. (2015). On tasks and cognitive constructs for the bilingual (non-)advantage. *Cortex, 73*, 347–348. https://doi.org/10.1016/j.cortex.2015.07.017.
Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377.
Gullifer, J. W., & Titone, D. (2018). *Compute language entropy with {languageEntropy}*. Retrieved from https://github.com/jasongullifer/languageEntropy.
Gullifer, J. W., & Titone, D. (2019). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition, 23*(2), 1–12. https://doi.org/10.1017/S1366728919000026.
Haaf, J. M., & Rouder, J. N. (2019). Some do and some don't? Accounting for variability of individual difference structures. *Psychonomic Bulletin & Review, 26*(3), 772–789. https://doi.org/10.3758/s13423-018-1522-x.
Hartanto, A., & Yang, H. (2016). Disparate bilingual experiences modulate task-switching advantages: A diffusion-model analysis of the effects of interactional context on switch costs. *Cognition, 150*, 10–19. https://doi.org/10.1016/j.cognition.2016.01.016.
Hartanto, A., & Yang, H. (2020). The role of bilingual interactional contexts in predicting interindividual variability in executive functions: A latent variable analysis. *Journal of Experimental Psychology: General, 149*(4), 609–633. https://doi.org/10.1037/xge0000672.
Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*, 1166–1186. https://doi.org/10.3758/s13428-017-0935-1.
Henrard, S., & Van Daele, A. (2017). Different bilingual experiences might modulate executive tasks advantages: Comparative analysis between monolinguals, translators, and interpreters. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01870.
Jaekel, N., Jaekel, J., Willard, J., & Leyendecker, B. (2019). No evidence for effects of Turkish immigrant children's bilingualism on executive functions. *PLoS One, 14*(1), e0209981. https://doi.org/10.1371/journal.pone.0209981.
Jylkkä, J., Soveri, A., Wahlström, J., Lehtonen, M., Rodríguez-Fornells, A., & Laine, M. (2017). Relationship between language switching experience and executive functions in bilinguals: An internet-based study. *Journal of Cognitive Psychology, 29*(4), 404–419. https://doi.org/10.1080/20445911.2017.1282489.
Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin, 144*(11), 1147–1185. https://doi.org/10.1037/bul0000160.
Kline, R. B. (1998). *Principles and practice of structural equation modeling.* pp. xiv, 354Guilford Press.
Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press pp. xvii, 534.
Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science, 23*(3), 159–163. https://doi.org/10.1177/0963721414528511.
Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0.
Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition, 17*(3), 673–680. https://doi.org/10.1017/S1366728913000606.
Linck, J. A. (2015). Methods matter for critical reviews too. *Cortex, 73*, 354–355. https://doi.org/10.1016/j.cortex.2015.07.011.
Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. *Inhibitory processes in attention, memory, and language* (pp. 189–239). Academic Press.
Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review, 91*(3), 295–327. https://doi.org/10.1037/0033-295X.91.3.295.
Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067).
Marzecová, A. (2015). Bilingual advantages in executive control – A Loch Ness Monster case or an instance of neural plasticity? *Cortex, 73*, 364–366. https://doi.org/10.1016/j.cortex.2015.07.005.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734.
van den Noort, M., Struys, E., Bosch, P., Jaswetz, L., Perriard, B., Yeo, S., Barisch, P., Vermeire, K., Lee, S.-H., & Lim, S. (2019). Does the bilingual advantage in cognitive control exist and if so, what are its modulating factors? A systematic review. *Behavioral Science, 9*(3), 27. https://doi.org/10.3390/bs9030027.
Ooi, S. H., Goh, W. D., Sorace, A., & Bak, T. H. (2018). From bilingualism to bilingualisms: Bilingual experience in Edinburgh and Singapore affects attentional control differently. *Bilingualism: Language and Cognition, 21*(4), 867–879. https://doi.org/10.1017/S1366728918000020.
Paap, K. (2019). The bilingual advantage debate: Quantity and quality of the evidence. In J. W. Schwieter, & M. Paradis (Eds.). *The handbook of the neuroscience of multilingualism* (pp. 701–735). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119387725.ch34.
Paap, K. R. (2014). The role of componential analysis, categorical hypothesising, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *Journal of Cognitive Psychology, 26*(3), 242–255. https://doi.org/10.1080/20445911.2014.891597.
Paap, K. R., Anders-Jefferson, R., Zimiga, B., Mason, L., & Mikulinsky, R. (2020). Interference scores have inadequate concurrent and convergent validity: Should we stop using the flanker, Simon, and spatial Stroop tasks? *Cognitive Research: Principles and Implications, 5*(1), 7. https://doi.org/10.1186/s41235-020-0207-y.
Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*(2), 232–258. https://doi.org/10.1016/j.cogpsych.2012.12.002.
Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex, 69*, 265–278. https://doi.org/10.1016/j.cortex.2015.04.014.
Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods, 274*, 81–93. https://doi.org/10.1016/j.jneumeth.2016.10.002.
Parsons, S., Kruijt, A.-W., & Fox, E. (2018). Psychological science needs a standard practice of reporting the reliability of cognitive behavioural measurements. Preprint *PsyArXiv*. https://doi.org/10.31234/osf.io/6ka9z.
Pot, A., Keijzer, M., & de Bot, K. (2018). Intensity of multilingual language use predicts cognitive performance in some multilingual older adults. *Brain Sciences, 8*(5), 92. https://doi.org/10.3390/brainsci8050092.
R Core Team (2019). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/

xlm0000450.
Rodriguez-Fornells, A., Krämer, U. M., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00388.
Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. http://www.jstatsoft.org/v48/i02/.
Rouder, J., Kumar, A., & Haaf, J. M. (2019). Why most studies of individual differences with inhibition tasks are bound to fail. Preprint *PsyArXiv*. https://doi.org/10.31234/osf.io/3cjr5.
Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review, 26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y.
Samuel, S., Roehr-Brackin, K., Pak, H., & Kim, H. (2018). Cultural effects rather than a bilingual advantage in cognition: A review and an empirical study. *Cognitive Science, 42*(7), 2313–2341. https://doi.org/10.1111/cogs.12672.
Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Lawrence Erlbaum Associates Publishers pp. xii, 498.
Ursache, A., & Noble, K. G. (2016). Neurocognitive development in socioeconomic context: Multiple mechanisms and implications for measuring socioeconomic status: SES and neurocognitive function. *Psychophysiology, 53*(1), 71–82. https://doi.org/10.1111/psyp.12547.
Valian, V. (2015a). Bilingualism and cognition. *Bilingualism: Language and Cognition, 18*(01), 3–24. https://doi.org/10.1017/S1366728914000522.
Valian, V. (2015b). Bilingualism and cognition: A focus on mechanisms. *Bilingualism: Language and Cognition, 18*(01), 47–50. https://doi.org/10.1017/S1366728914000698.
Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences, 12*(11), 418–424. https://doi.org/10.1016/j.tics.2008.07.005.
Verreyt, N., Woumans, E., Vandelanotte, D., Szmalec, A., & Duyck, W. (2016). The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition, 19*(1), 181–190. https://doi.org/10.1017/S1366728914000352.
Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100*(3), 426–432. https://doi.org/10.1037/a0022790.